DOCUMENT RESUME

ED 185 916 HE 012 598

AUTHOR Webster, David S.
TITLE How to Assess Quality in Master's Degree Programs--A New and Better Way.
PUB DATE 1 Mar 79
NOTE 12p.; Paper presented at a conference on assessing quality in master's degree programs (University of Maryland, College Park, MD, March 1, 1979). Best copy available.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS College Environment; College Faculty; *Departments; Educational Facilities; *Educational Quality; *Evaluation Criteria; Evaluation Methods; Grades (Scholastic); *Graduate Study; Higher Education; Job Placement; *Masters Programs; *Program Evaluation; Reputation; Scholarship; Standardized Tests; Student Characteristics; Student Employment; Student Experience; Student Teacher Relationship; Teacher Effectiveness; Teacher Guidance

ABSTRACT

Evaluation methods for master's degree programs should be substantially different from those currently used for doctoral degree programs. Three current methods, that of Beverley Hurlbert, that of Jack Gourman, and that using departmental reputations as the major criterion, are insufficient. Elements that should be included in the evaluation are the department itself, ancillary departments that could reasonably be expected to attract a substantial number of students from the department being rated, campus facilities used by the department, and the advantages and disadvantages of the surrounding community for the subject being studied. Department measures include: the faculty's scholarship; amount of time faculty devote to teaching, advising, guiding student research, and student contact; quality of students, measured by Graduate Record Examination scores, undergraduate grades weighted by college quality, and current jobs and cumulative experience; student participation in department and campus life; and graduate placement record. Campus-wide measures include strength and accessibility of related departments, and the quality of facilities a program draws upon most. Although the community environs should be considered, they should not be weighted too heavily. A list of references is included. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185916

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

David S. Webster

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

HOW TO ASSESS QUALITY IN MASTER'S DEGREE PROGRAMS --

A NEW AND BETTER WAY

An unpublished paper delivered at the conference on "How to Assess Quality in Master's Degree Programs," University of Maryland University College, March 1, 1979.

David S. Webster
Chancellor's Intern Fellow
UCLA
April, 1980.

HE 012 598

HOW TO ASSESS QUALITY IN MASTER'S PROGRAMS -- A NEW AND BETTER WAY

David S. Webster

University of California at Los Angeles

The best-known studies of the quality of graduate programs virtually all rate the quality of Ph.D. programs; these studies have severe shortcomings, even as ratings of doctoral programs, and when techniques are developed for rating master's degree programs, they will need to be very different from the methods widely used now. Let's look, briefly, at three currently used techniques of assessing the quality of graduate programs.

1. One method, proposed by anthropologist Beverley Hurlbert (4), uses exchange theory in rating anthropology departments. When departments hire each other's Ph.D.'s, she assumes that they are equal in status, but if, for example, the University of Florida hires several of Howard's Ph.D.'s, and Harvard does not reciprocate, she concludes that Harvard is higher in status; thus, by examining recent hiring patterns among universities she rank orders anthropology departments. This method has serious flaws, including the following:

a. it can measure only the strength of departments relative to each other, not the quality of each department.

b. it assumes that university hiring practices are entirely meritocratic.

c. it measures a department's strength by only those graduates who receive appointments in Ph.D.-granting institutions. An

economics department, for example, that produced 20 Ph.D.'s, of who two became assistant professors of economics at Ph.D.-granting universities and 18 became hoboes, would be rated higher than a department that turned out 20 Ph.D.s, all of who were hired for responsible jobs by major corporations and government agencies.

d. it rates departments solely by the first jobs of their graduates, and thus attributes too much importance to the quality of a program's students. Many studies by Alexander W. Astin (see, for example, 1) and others have shown that the academic achievement of college seniors is much more influenced by the ability of the students when they entered than by the institution's "quality," and if this finding holds true for graduate education, and if the best students get the best jobs, then the "exchange" method really measures the quality of a program's students more than the quality of the program.

2. Another method of rating the quality of graduate programs has been employed by Jack Gourman (3), and while Hurlbert's technique rates programs largely on one factor -- the quality of their students -- Gourman goes to the opposite extreme and considers a great number of things. In his "Method of Evaluation" section he says he has evaluated each institution's "auspices, control and organization," its "total education programs offered," its "number of non-departmental areas," age, admissions requirements, student enrollment, teaching staff, physical facilities, finances, curriculum and research.

But Gourman gives scarcely a clue to how he measured all these things, scarcely a clue to how important he considered each

of them (how, for example, do you weight a school's "teaching staff" against its "age," and does a school gain or lose points because it is old? Also, when Gourman assesses a college's "auspices, control and organization," what types of auspices and organization improve one's rating, and what types hurt it? Gourman is silent on these and many other methodological matters.

At any rate, Gourman attempts no ratings of graduate schools by field, except for law, medicine, and dentistry. One normally enters a graduate program to study a particular subject -- even if one's program is interdisciplinary, it is still normally based on a few related fields -- and Gourman's global ratings of universities tell little, except for the fields of law, medicine, and dentistry, about any particular field of study.

3. But in a sense we have been setting up straw men so far, for the two rating methods we have discussed have been used, so far as we know, by only a single author. By far the most widely used means of evaluating graduate programs has been to assess the research reputation of their faculty by asking others in their discipline -- department chairmen, senior scholars, junior scholars, or some combination -- to rate them. This technique has been used, in one form or another, in the two best-known ratings, those of Allan Cartter (2) and Roose and Andersen (6), and also in the study by Ladd and Lipset (5), whose preliminary findings were summarized in the Chronicle of Higher Education a few weeks ago.

These "reputation of faculty" studies have been the best-designed and most influential of the existing quality studies, but they measure, for the most part, only the quality of a department's

faculty and neglect most other aspects of it. Cartter, for his part, did include, at the end of his study, a section in which he rated a university's library resources and the number of its graduate students who had won national fellowships, but he made no attempt to integrate these measures with his ratings of faculty to form one, overall measure of departmental quality. And CArtter and Roose-Andersen both tried to rate not only the quality of a department's faculty but the overall effectiveness of the program, yet the two scores correlate so nearly perfectly -- for the 36 disciplines studied by Roose-Andersen the correlations were all either .98 or .99, usually .99 -- that what was being measured was clearly the same thing.

What, then, is to be done? To measure the quality of a school's graduate program in any discipline we must measure not only the reputation of its faculty, as most of the ratings do, not just the ability of its students, as Hurlbert's study does, and certainly not the entire institution, as Gourman says he does[1] -- but the department itself, ancillary departments whose offerings could reasonably be expected to attract a substantial number of students from the department being rated, campus facilities used by the department, and the advantages and disadvantages of the surrounding community for the subject being studied.

---

[1]For example, the research potential of Harvard University's forest, located in central Massachusetts, probably does not vitally affect its students studying Comparative Literature in Cambridge, and the fact that New York University, according to the Roose-Andersen survey, had the best art history department in the country, might have been more relevant to philosophy students wishing to specialize

What follows is an outline of a rating system according to the principles we have described. We will classify our measures into those dealing with the department being rated, those involved with things outside the department but on the campus, and those concerned with the community surrounding the campus.

1. Departmental Measures

a. the faculty's scholarship -- the research abilities of the faculty, rated -- as in many existing assessments -- by other scholars in the same discipline.

b. amount of time faculty devote to teaching, advising, guiding student research, and student contact in general -- in many highly rated departments the most visible people, the ones upon whom the department's reputation is largely based, do less teaching and advising, and have less contact with students, than less well-known people. Thus a program's reputation for teaching effectiveness may be based, paradoxically, upon the reputations of those who do the least teaching. Therefore, in rating programs we should consider not only their professors' research productivity but the contribution they make towards teaching and advising students, and a professor who is often on leave and teaches a light schedule when he is in residence should not "count" as much, in measuring a program's teaching effectiveness, as one who does a full share of teaching and advising.

c. quality of students -- this would be measured by GRE aptitude and achievement scores, undergraduate grades weighted by quality of college attended, and -- in the case of some professional

Master's programs -- students' current jobs and cumulative experience in their field.

d. student participation in department and campus life -- these could be measured both by the percentage of full-time students, of students who served as teaching or research assistants, and of students who lived on or very near campus, and also by the absolute number of students in a department who were fully involved in department and campus life. All things being equal, programs with a high percentage of such fully involved students would be rated higher than others; however, having a "critical mass" of students for others to interact with is important, too, and a graduate program with, say, 100 students, 40 of who were full-timers, would be rated higher than a program with 10 students, four of who were full-timers.

e. placement record of graduates -- here I would consider not only initial academic appointments, as Hulbert does, but how well the program served to advance all its students' careers. In the case of a Master's program, for example, I would consider the percentage of its graduates who went on for doctorates, the percentage of those who actually completed their doctorates, the quality of the Ph.D-granting institution they went to, and also the kinds of jobs the terminal Master's students got, their starting salaries, and -- for those who were employed before beginning their program -- their salaries and responsibilities, immediately after earning their Master's, compared to what these were just before they entered the program.

2. Campus-wide Measures

a. strength of related departments -- in rating a Master's program in Special Education, for example, I would consider the quality of the campus' psychology department, especially the number and quality of the course offerings and the quality of those faculty most valuable to a student studying Special Education, and also the accessibility of these psychology courses and faculty to Special Education students.

b. quality of campus facilities a program draws upon -- instead of using global measures, as does Gourman, and assigning a rating to an entire institution, I would attempt, for each program, to rate those campus academic facilities it uses most. In the case of a Master's program in English, for example, I would rate the central university library and the English or humanities library, if any, but not the computer network; for a program in art, the campus' studios, equipment, and museum collections, if any; for a program in sociology, both the campus' library and its computer network. Though we cannot predict with certainty, of course, what campus resources a program's students will use -- there will always be students in Special Education who have little or no contact with the campus' psychology department, and some students in English who wish to do computer-assisted studies -- selecting the campus facilities students in a given program are likely to use most and measuring the quality of those is at least a step in the right direction.

3. Surrounding Community

Finally, we should consider, in assessing any program's quality, the campus' wider environs, and we should rate both the community's intellectual ambience in general and what it offers for the particular field being studied. For example, Harvard and UC Berkeley might well earn a higher rating for their neighborhoods, with their many bookstores and cultural opportunities, than schools less favorably situated.

Turning to ratings of the campus' environment in terms of a particular field, we would argue that New York City is a good environment to study most periods of the history of western art, Washington, D.C. is a good place to study many aspects of American government, and schools very near an ocean -- for example, the University of Miami in Coral GAbles and UC San Diego -- are good places to study marine biology. Programs located in such environments would be rated higher, all things being equal, than similar programs less favorably situated.

In suggesting that a great many factors be considered in rating graduate programs, I am aware that I have suggested rating some things -- for example, the quality of a campus' surrounding community -- that are extremely hard to measure. And I know that so far as campus environments are concerned, as with many other things, one scholar's Paradise is another's Purgatory, and that some people, particularly in the late 1960s, found the Berkeley environment I have just touted an exceptionally unpleasant place to be.

Furthermore, though I have praised certain types of environ-

ments as ideal settings for certain disciplines, it is clear that the finest departments in the country are by no means always, or even often, in the settings that would seem to be the most hospitable. For example, according to the 1970 Roose-Andersen ratings, the finest Spanish department in the country was not anywhere in the Sunbelt but rather at Harvard, and the third and fourth best Spanish departments were in decidedly un-Hispanic Madison, Wisconsin and Philadelphia. Similarly, the best German department in the country was not in Milwaukee but at Berkeley, and the fourth best philosophy department was in Pittsburgh, a town that has reminded few people of Socrates' Athens.

On the other hand, though our nation's capital would seem to be a good place to study American government, of the nation's 54 best political science departments, again according to Roose-Andersen, not one was located in Washington. In short, we must be careful not to weigh a program's environs too much. How much, though, is enough? How much importance should we assign to each of our different measures? If we are rating the campus' surroundings, the school's library, and the quality of, say, the history faculty, should we weight each of these equally, and, if not, how should we assign weights? These are serious problems, to which there are no simple solutions. Still and all, by rating graduate programs on a large number of factors, and by weighting these factors appropriately -- difficult as this may be -- we will come much closer to measuring the true quality of graduate programs than we do with the one-dimensional measure -- the measure of the faculty's research reputation -- that is commonly used today.

REFERENCES

1. Astin, Alexander W. "Undergraduate Achievement and Institutional 'Excellence,'" Science, 161 (3842), 1968, pp. 661-668.

2. Cartter, Allan M. An Assessment of Quality in Graduate Education (Washington, D.C.: American Council on Education), 1970.

3. Gourman, Jack. The Gourman Report: a Rating of American and International Universities (Los Angeles: National Education Standards), 1977.

4. Hurlbert, Beverley McElligott. "Status and Exchange in the Profession of Anthropology," American Anthropologist, (78), 1976, pp. 272-284.

5. Ladd, Everett Carl and Seymour Martin Lipset. "How Professors Rated Faculties in 19 Fields," Chronicle of Higher Education, January 15, 1979, p. 6.

6. Roose, Kenneth D. and Charles J. Andersen, A Rating of Graduate Programs (Washington, D.C.: American Council on Education), 1970.